Некоторые решения задач контрольных работ 3 (Б.Демешев)

2018-2019

1. У Билла Гейтса было 29 наблюдений, так как в матрице $X'X$ верхний левый элемент — это число наблюдений. Средние равны $\bar{X} = 0/29 = 0, \bar{W} = 0/29 = 0$. Подставляем средние значения регрессоров в уравнение регрессии и получаем средние значение предсказываемой переменной $\bar{Y} = 4$.
2. Находим ожидания: $\mathbb{E}(X_i) = a$, $\mathbb{E}(|X_i|) = 1.5a \cdot 0.75 + 4a \cdot 0.25 = 2.125a$.
   Отсюда получаем моментные условия. Первое: $g_1(X_i, a) = X_i - a$, следовательно, $\bar{g}_1 = \bar{X} - a$. Второй: $g_2(X_i, a) = |X_i| - a$, следовательно, $\bar{g}_2 = \sum |X_i|/n - 2.125a$.
   Если нужна оценка метода моментов, то получаем, что $\bar{X} - \hat{a} = 0$, следовательно, $\hat{a}_{MM} = \bar{X}$.
   Если нужна оценка обобщённого метода моментов, то нужно минимизировать функцию:
   $$\begin{pmatrix} \bar{X} - a & \sum |X_i|/n - 2.125a \end{pmatrix} \cdot \begin{pmatrix} 3 & 0 \\ 0 & 64 \end{pmatrix} \cdot \begin{pmatrix} \bar{X} - a \\ \sum |X_i|/n - 2.125a \end{pmatrix}$$
   В силу нулевых внедиагональных весов задача упрощается до
   $$Q(a) = 3(\bar{X} - a)^2 + 64(\sum |X_i|/n - 2.125a)^2 \to \min_a$$

2017-2018

1. Найдём оценку ковариационную матрицу оценок коэффицентов:
   $$\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = 0.04 \begin{pmatrix} 1/14 & 0 \\ 0 & 1/77 \end{pmatrix},$$

   дисперсию ошибки прогноза:
   $$\begin{aligned}
   \widehat{\mathrm{Var}}(y_i - \hat{y}_f | X) &= \widehat{\mathrm{Var}}(\beta_z z_i + \beta_w w_i + \varepsilon_i - \hat{y}_f | X) \\
   &= \widehat{\mathrm{Var}}\,(\varepsilon_i | X) + \widehat{\mathrm{Var}}\left(-2\hat{\beta}_z + 5\hat{\beta}_w | X\right) - 2\widehat{\mathrm{Cov}}\,(\varepsilon_i, \hat{y}_f) \\
   &= 0.04 + 4\widehat{\mathrm{Var}}\left(\hat{\beta}_z | X\right) + 25\widehat{\mathrm{Var}}\left(\hat{\beta}_w | X\right) - 2 \cdot (-2) \cdot 5\widehat{\mathrm{Cov}}\left(\hat{\beta}_z, \hat{\beta}_w\right) + 0 \\
   &= 0.04 + 4 \cdot 0.04 \cdot \frac{1}{14} + 25 \cdot 0.04 \cdot \frac{1}{77} + 0 \\
   &\approx 0.0644,
   \end{aligned}$$

   и сам прогноз:
   $$\hat{y}_f = 0.2 \cdot (-2) + 0.3 \cdot 5 = 1.1.$$

   Теперь можно выписать доверительный интервал, $t_{0.975,4} = 2.78$:
   $$\left[1.1 - 2.78\sqrt{0.0644}; 1.1 + 2.78\sqrt{0.0644}\right]$$

2. Функция правдоподобия в неограниченной модели ($\sigma_1^2 \neq \sigma_2^2$) имеет вид:

$$L\left(\sigma_1^2, \sigma_2^2\right) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma_1^2}} \cdot \prod_{i=m+1}^{n} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma_2^2}}$$

Выпишем также логарифмическую функцию правдоподобия для неограниченной модели и найдём оценки $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$:

$$\begin{aligned}
\ell\left(\sigma_1^2, \sigma_2^2\right) = & -\frac{m}{2}\ln 2\pi - \frac{m}{2}\ln\sigma_1^2 - \frac{1}{2\sigma_1^2}\sum_{i=1}^{m}\left(y_i - x_i'\beta\right)^2 \\
& -\frac{n-m}{2}\ln 2\pi - \frac{n-m}{2}\ln\sigma_2^2 - \frac{1}{2\sigma_2^2}\sum_{i=m+1}^{n}\left(y_i - x_i'\beta\right)^2 \to \max_{\sigma_1^2, \sigma_2^2} \\
\frac{\partial \ell}{\partial \sigma_1^2} = & -\frac{m}{2\sigma_1^2} + \frac{1}{2\left(\sigma_1^2\right)^2}\sum_{i=1}^{m}\left(y_i - x_i'\beta\right)^2 \Bigg|_{\sigma_1^2 = \hat{\sigma}_1^2} = 0 \\
\frac{\partial \ell}{\partial \sigma_2^2} = & -\frac{n-m}{2\sigma_2^2} + \frac{1}{2\left(\sigma_2^2\right)^2}\sum_{i=m+1}^{n}\left(y_i - x_i'\beta\right)^2 \Bigg|_{\sigma_2^2 = \hat{\sigma}_2^2} = 0 \\
\hat{\sigma}_1^2 = & \frac{\sum_{i=1}^{m}\left(y_i - x_i'\beta\right)^2}{m} \\
\hat{\sigma}_2^2 = & \frac{\sum_{i=m+1}^{n}\left(y_i - x_i'\beta\right)^2}{n-m}
\end{aligned}$$

Тогда логарифмическая функция правдоподобия примет вид:

$$\ell_{UR}\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2\right) = -\frac{n}{2}\ln 2\pi - \frac{n}{2} - \frac{m}{2}\ln\frac{\sum_{i=1}^{m}\left(y_i - x_i'\beta\right)^2}{m} - \frac{n-m}{2}\ln\frac{\sum_{i=m+1}^{n}\left(y_i - x_i'\beta\right)^2}{n-m}$$

Проделаем то же самое для ограниченной модели ($\sigma_1^2 = \sigma_2^2 = \sigma_0^2$):

$$\begin{aligned}
L\left(\sigma_0^2\right) = & \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma_0^2}} \\
\ell\left(\sigma_0^2\right) = & -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma_0^2 - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2 \to \max_{\sigma_0^2} \\
\frac{\partial \ell}{\partial \sigma_0^2} = & -\frac{n}{2\sigma_0^2} + \frac{1}{2\left(\sigma_0^2\right)^2}\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2 \Bigg|_{\sigma_0^2 = \hat{\sigma}_0^2} = 0 \\
\hat{\sigma}_0^2 = & \frac{\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2}{n} \\
\ell_R\left(\hat{\sigma}_0^2\right) = & -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\frac{\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2}{n} - \frac{n}{2}
\end{aligned}$$

Осталось выписать формулу статистики LR-теста:

$$\begin{aligned}
LR & = 2(\ell_{UR} - \ell_R) \\
& = 2\left(-\frac{m}{2}\ln\frac{\sum_{i=1}^{m}\left(y_i - x_i'\beta\right)^2}{m} - \frac{n-m}{2}\ln\frac{\sum_{i=m+1}^{n}\left(y_i - x_i'\beta\right)^2}{n-m} + \frac{n}{2}\ln\frac{\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2}{n}\right)
\end{aligned}$$

3. Наиболее эффективная оценка коэффициента β может быть получена с помощью взвешенного МНК. Для этого необходимо оценить регрессию

$$\frac{y_i}{\sqrt{i}} = \beta\frac{x_i}{\sqrt{i}} + \frac{\varepsilon_i}{\sqrt{i}}.$$

Переобозначив $\frac{y_i}{\sqrt{i}}$ за y_i^*, $\frac{x_i}{\sqrt{i}}$ за x_i^* и $\frac{\varepsilon_i}{\sqrt{i}}$ за ε_i^*, получим регрессию

$$y_i^* = \beta x_i^* + \varepsilon_i^*,$$

применив к которой обычный МНК, получим эффективную оценку $\hat{\beta}_{WLS}$ вида

$$\hat{\beta}_{WLS} = \frac{\sum_{i=1}^n x_i^* y_i^*}{\sum_{i=1}^n (x_i^*)^2} = \frac{\sum_{i=1}^n \frac{x_i}{\sqrt{i}} \cdot \frac{y_i}{\sqrt{i}}}{\sum_{i=1}^n \frac{x_i^2}{i}}.$$

4. а) $\beta_0 : \frac{2-0}{\sqrt{0.25}} = 4 > 2 \Rightarrow$ гипотеза о незначимости коэффициента отвергается
$\beta_1 : \frac{3-0}{\sqrt{0.16}} = 7.5 > 2 \Rightarrow$ гипотеза о незначимости коэффициента отвергается

б)

$$\begin{aligned}\frac{\partial\hat{\mathbb{P}}(Y_i = 1)}{\partial X_i} &= F'\left(\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i\right) \cdot \hat{\beta}_1 \\ &= F\left(\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i\right)\left(1 - F\left(\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i\right)\right) \cdot \hat{\beta}_1\Big|_{X_i=0.5} \\ &= \frac{e^{3.5}}{1+e^{3.5}}\left(1 - \frac{e^{3.5}}{1+e^{3.5}}\right) \cdot 3 \\ &\approx 0.085\end{aligned}$$

в) Предельный эффект максимален в точке, где наклон касательной к логистической функции самый крутой, то есть в нуле:

$$2 + 3x = 0$$

Формально получается $x = -\frac{2}{3}$, однако в данной задаче x неотрицательная величина, поэтому оптимум оказывается в точке $x = 0$.

2016-2017

1. Второе уравнение оценили для корректного построения доверительных интервалов и проверки гипотез о коэффициентах в условиях гетероскедастичности, для получения более эффективных оценок, $\mathrm{Var}(u_i) = \sigma^2 X_i$.

   После применения взвешенного МНК оба коэффициента значимы.

2. Собственные значения ковариационной матрицы: $\lambda_1 = 1.3$, $\lambda_2 = 1$, $\lambda_3 = 0.7$.

   Веса для стандартизированных переменных в первой главной компоненте: $\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{pmatrix}$

   Доля дисперсии: $\frac{13}{30}$

3. $$F = \frac{(RSS_R - RSS_{UR})/3}{RSS_{UR}/(n - k_{UR})} = \frac{(17-10)/3}{10/(25-6)} = 4.4(3)$$